



Table of Contents

Preface

Summary Article

Individual Contributions

Statistics as the information science

Statistical issues for databases, the internet, and experimental data

Mathematics in image processing, computer graphics, and computer vision

Future challenges in analysis

Getting inspiration from electrical engineering and computer graphics to develop interesting new mathematics

Research opportunities in nonlinear partial differential equations

Risk assessment for the solutions of partial differential equations

Discrete mathematics for information technology

Random matrix theory, quantum physics, and analytic number theory

Mathematics in materials science

Mathematical biology: analysis at multiple scales

Number Theory and its Connections to Geometry and Analysis

Revealing hidden values: inverse problems in science and industry

Complex stochastic models for perception and inference

Model theory and tame mathematics

Beyond flatland: the future of space

## Revealing Hidden Values: Inverse Problems in Science and Industry

J. MCLAUGHLIN AND W. SYMES

The last century saw a vast increase in the degree to which key processes in science, engineering, and business are described by mathematical models. Of course this is a 350-year-old story in the physical sciences going back to Kepler's mathematical description of the motion of the planets and Newton's Law for the gravitational force one body exerts on another. Much newer are discoveries of mathematical models that capture the essence of many important phenomena in biosciences, social sciences, and [even] finance. Very often, however, the general structure of the model is understood but specific features or characteristic numbers are not. Examples of such key parameters, without which model-based predictions are impossible, but which are equally impossible to measure directly, are: the volatility of a stock price, needed to optimally price a European stock option; the reaction rates in cell cycle models; the conductivity or density of minerals and ores in the Earth's mantle, used then to identify the location of those minerals; the structural changes due to cracks and flaws in materials, used to identify the location of the flaws; or the change in stiffness in biological tissue due to cancerous growths, used then to identify those growths. In order to calibrate mathematical models and render them useful to practitioners of each relevant discipline, these hidden values must be extracted from measurements via the models themselves.

The advent of large scale computing over the last few decades has led to an explosion of opportunities to exploit this use of mathematical models. The technique is often called "inverse," as the models are used "backwards" to infer their underlying parameters from measurements, rather than to predict the measurements from the parameters. In Newton's law, for example, the speed would not be found knowing the gravitational acceleration constant but rather the gravitational acceleration constant is determined from velocity measurements.

One might call these methods "model rich" data mining methods since (often extremely large) data sets are analyzed with the aid of mathematical models. Note that computational power is not enough: without deep understanding of natural and mathematical structure underlying this wonderful new data, the rapid advance of computer capability cannot achieve its potential to be used to extract technological and societal implications! This fact is often lost in popular discussion of the digital revolution. In this document we will overview a few of the more exciting recent developments in "model rich data mining", also known as inverse problems, observing in each case mathematical challenges and extra-mathematical implications.

One of the most exciting of these areas to emerge in recent years applies to the computer itself, or rather to the Internet. The Net is a vast and rapidly changing collection of connections (edges) and computers (nodes): in fact it is so large, and dynamic, that a detailed description of it as a network, in the traditional mathematical sense, is very unwieldy and of little practical value. However at a coarse "spatial" scale \_ measured in connectivity, not distance \_ and over short time-windows it has statistically stable behavior. This apparent stability opens up the possibility of a description of net traffic that does not depend on detailed tracking of each and every packet. Instead, researchers divide the Net into regions and seek the rate of packets passing between contiguous regions. Assuming a statistical model (for example independent Poisson processes) for each source-destination pair, it's possible to estimate the parameters of the processes from either data on contiguous-region traffic recorded by routers, or even from data on end-to-end traffic for only a few of the myriad source-destination pairs. This is a

and time

Mathematics in  
molecular biology  
and medicine

The year 2000 in  
geometry and  
topology

Computations and  
numerical  
simulations

Numbers, insights  
and pictures: using  
mathematics and  
computing to  
understand  
mathematical  
models

List of Contributors  
with Affiliations

problem of "inverse" type, called "network tomography." Having estimated the statistics of contiguous region traffic, it is possible to reconstruct the statistics of traffic between any two regions within the Net, hence obtain economically valuable information about who is messaging whom and how much, what changes in traffic result from changes in connectivity, etc. A successful probabilistic model will make management of the Net and its rapid growth much more tractable; with the economic value of Net activity soaring past one trillion dollars in the near future, this prospect has attracted a great deal of attention. A worldwide cohort of applied mathematicians, statisticians, and signal processing engineers are attacking this modeling problem. Significant mathematical challenges posed by the approach described here include devising methods for subdividing the Net to maximize the effectiveness of tomography, understanding what statistical models of contiguous region traffic are appropriate, and developing algorithms to estimate the parameters of fairly general statistical traffic models.

Another set of intriguing and only recently recognized inverse problems arises in finance. Asset pricing for various types of traded options engages a substantial segment of the securities industry, with cash flows in the hundreds of billions daily. Introduction of stochastic partial differential equation models for price-volatility relations has had an enormous impact on this business (and led to a Nobel Prize in Economics several years ago). The volatilities that play a central role in these models can vary over time and are not directly measurable. Given estimates of volatilities, the models predict prices via various numerical methods, e.g., diffusion approximation or Monte Carlo simulation. The essential task is to calibrate the models so that they correctly predict a set of past option prices. Calibrated models yield accurate future price predictions, and so lead to successful portfolio management and, ultimately, a more efficient economy. Estimating volatilities is not a simple task. Major mathematical difficulties flow from the stochastic properties of the data sets which are the stock market prices, the subsequent stochastic properties of the options prices, and the need to determine the volatilities in real time. Progress in resolving these difficulties is essential to realizing the promise of option pricing models, and can be made by the combined efforts of a growing group of mathematicians, economists, and statisticians.

Modern science and technology pose a variety of similar problems. Biomedicine relies on indirect inference for real-time information about the body. X-ray CAT and MRI, the most extensively developed medical imaging technologies, are literally lifesavers. Critical information still lies beyond that extractable with current methods, however, and even these techniques still pose interesting mathematical challenges with potential payoff in better resolution and so better diagnosis and national health. Two other techniques that rely on measurements of vibrations of tissue are just beginning to have careful scrutiny. The first is diffuse optical tomography, in which tissue is illuminated with visible spectrum light and the transmission characteristics are interpreted digitally for tissue health--a sort of high-tech "visual examination." Since visible light scatters strongly from tissue, the information is highly diffused and difficult to localize. However, sound waves at ultrasonic frequencies compress and expand the tissue microscopically, and modulate the scattering of visible light so that tissue anomalies appear much easier to localize. The second vibration technique goes by the name elastography. It is inspired by the familiar doctor's examination where palpation of accessible tissue can identify stiffer than normal regions, providing an alert to possible problems. To extend this idea to the interior tissue, low frequency waves, which the body tolerates well, create excitations whose amplitudes are measured by ultrasound techniques. The key feature is that stiff regions show lower amplitudes than normal tissue. Algorithm development for ultrasonically modulated diffuse tomography and for the low frequency elastography technique is just beginning, and many essentially mathematical questions must be addressed to bring these techniques to practically useful states.

Earth sciences, broadly construed, continue to be a generator of many compelling inverse problems. All of our knowledge of the Earth's interior is indirectly derived from measurements, as is a great deal of what we know about the surface and the atmosphere. Besides well-known and economically important topics, such as reflection seismology in oil exploration and LandSat multispectral imaging of vegetation and terrain, a wide variety of other technological imperatives drive work on indirect inference--the Army wants to find tanks under foliage using low frequency radar, the Navy wants to locate mines in shallow sediments near beaches from sonar returns, environmental engineers need precise locations of aquicludes impeding the motion of groundwater contaminant plumes from radar or high resolution seismic measurements, airframe manufactures need better means of discriminating eddy current anomalies due to cracks and fractures from measurement noise, etc. Some of the most exciting opportunities will arise over the next few years as a flood of new Earth monitoring data becomes available, for example, the terabyte per month or so of teleseismic data expected from the NSF EarthScope USArray seismic network now being installed and the terabyte of multispectral reflectometry data which will shortly begin to flow from the NASA Earth Observing System satellite. By their sheer size, these data sets will pose new challenges for image extraction and integrated interpretation. Many of these

challenges are essentially mathematical in nature.

All of these "inverse" problems share several important characteristics. Indeed, their solutions are important to society and the problems are hard. Somewhat less obvious but typical and very important features include nonlinearity, very large and information-rich data sets, and intrinsic uncertainty. The first two features underline the necessarily interdisciplinary nature of successful work on these problems: mathematicians must work together with scientists and engineers to ensure that the models are complicated enough, but not more complicated than is needed, to describe the natural phenomena, and to focus on the essential qualities of data. This interdisciplinary approach is necessary to ensure successful exploitation of the exciting opportunities for higher fidelity to natural phenomena opened up by the advance of computing power. Both nonlinearity and large numbers of model and data parameters are very often the result of dropping the oversimplifications which have made models mathematically and computationally tractable in the past. The third feature, intrinsic uncertainty in the results of inverse inference, can arise because of (inevitable) data and numerical errors, but also because small changes in very accurate data may correspond to large changes in properties being estimated. This phenomenon is known as illposedness; nonlinear relations and large data volumes make handling it more complicated. In terms of technical difficulties, these problems are worlds apart from estimating the gravitational constant from observations of falling cannon balls although the latter was certainly an impressive feat for its time. Substantial success on the serious mathematical issues of nonlinearity, illposedness, and very large (essentially infinite) dimension has taken place in the last twenty years and now advances rapidly, mostly due to substantial increase in understanding in nonlinear analysis. This is a vast area where a rich set of results can be expected and the payoff in problem solution methods is very important. Accompanying algorithms that effectively quantify uncertainties, deal with illposedness, and fully take the nonlinear model into account are needed.

Last Modified:  
Oct 17, 2001

[Previous page](#) | [Top of this page](#) | [Next page](#)

[Policies and  
Important Links](#)

| [Privacy](#)

| [FOIA](#)

| [Help](#)

| [Contact  
NSF](#)

| [Contact Web  
Master](#)

| [SiteMap](#)



The National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230,  
USA  
Tel: (703) 292-5111, FIRS: (800) 877-8339 | TDD: (800) 281-8749

Last Updated:  
10/17/01  
[Text Only](#)